Supplementary Materials for

Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging

Hyun-Woo Rhee, Peng Zou, Namrata D. Udeshi, Jeffrey D. Martell, Vamsi K. Mootha, Steven A. Carr, and Alice Y. Ting*

*To whom correspondence should be addressed. Email: ating@mit.edu

This PDF file includes:

Materials and Methods
Figures S1 to S11
References (*27-53*)

Other Supplementary Materials for this manuscript:

Supplementary Spreadsheets 1 and 2 containing Tables S1 to S9

Materials and Methods

Characterization of APEX-mediated biotinylation by fluorescence imaging (related to Figs. 1B, S1, S2, S4, and S7)

HEK293T or COS7 cells were cultured in MEM supplemented with 10% fetal bovine serum at 37 °C under 5% CO_2. Cells were transfected with the APEX fusion construct of interest using Lipofectamine 2000 (Invitrogen). Typically, 0.2 μg plasmid and 0.5 μL (for COS7 cells) or 1.0 μL (for HEK293T cells) Lipofectamine in MEM (without serum) was used for a 7×7 mm coverslip of cells 60-80% confluence. 4-6 hours after transfection, the cell culture medium was changed back to MEM supplemented with 10% fetal bovine serum. 18-24 hours after media change, biotin-phenol (final concentration C_f 500 μM) was added in the cell culture media and the cells were incubated for 30 minutes at 37 °C. Then H_2O_2 (C_f 1 mM) was added for 1 minute to initiate biotinylation. To halt the reaction, the cells were washed three times with a "quencher solution" consisting of 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in DPBS (Dulbecco's Phosphate Buffered Saline). Thereafter, the cells were fixed in 4% paraformaldehyde in PBS at room temperature for 15 minutes. Cells were then washed three times with PBS and permeabilized with pre-chilled methanol at -20 ºC for 5 minutes.

To detect biotinylated proteins, the fixed HEK293T or COS7 cells were stained with 50 nM neutravidin-AF (AlexaFluor) or streptavidin-AF, after blocking overnight in 1% casein in PBS ("blocking buffer") at 4 ºC. Neutravidin and streptavidin fluorophore conjugates, which could be used interchangeably without a difference in performance, were prepared by coupling neutravidin protein from Invitrogen or homemade streptavidin protein (*27*) with AF647-NHS or AF568-NHS esters (Invitrogen) following the manufacturer's instructions.

To detect APEX expression, we also stained the cells with anti-Flag (Sigma; 1:300 dilution), anti-V5 (Invitrogen; 1:2000 dilution), or anti-myc (Millipore; 1:1000 dilution) mouse antibodies in blocking buffer for 1 hour at room temperature. Samples were then washed 4 × 5 minutes with PBS and incubated with secondary goat anti-mouse-AF antibody (Invitrogen; 1:1000 dilution) in blocking buffer for 1 hour at room temperature. Samples were washed 4 × 5 minutes with PBS, then maintained in DPBS at room temperature for imaging.

Imaging was performed in confocal mode using a Zeiss AxioObserver.Z1 microscope equipped with a Yokogawa spinning disk confocal head and a Cascade II:512 camera. The confocal head contained a Quad-band notch dichroic mirror (405/488/568/647 nm). Samples were excited by solid state 491 (~20 mW), 561 (~20 mW), or 640 nm (~5 mW) lasers. Images were acquired using Slidebook 5.0 software (Intelligent Imaging Innovations), through a 63× oil-immersion objective for YFP/AF488 (528/38 emission filter), AF568 (617/73 emission filter), AF647 (700/75 emission filter), and differential interference contrast (DIC) channels. Acquisition times ranged from 20 to 500 milliseconds. Imaging conditions and intensity scales were matched for each dataset presented together.

In Fig. S1, mitochondria were stained by incubating cells with 500 nM Mitotracker Deep Red (Invitrogen) in media for 30 minutes at 37 ºC, prior to cell fixation.

Characterization of APEX-mediated biotinylation by streptavidin blotting of whole cell lysate (related to Figs. 1D, S1, S3, and S5)

HEK293T cells transfected and labeled as above were washed 3 times with quencher solution, then scraped and pelleted by centrifugation at 200 rpm for 5 minutes. The pellet was stored at -80 °C, then lysed with RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1× protease cocktail (Sigma Aldrich, catalog no. P8849), 1 mM PMSF (phenylmethylsulfonyl fluoride), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox) for 1 minute at 4 °C. The cell pellet was resuspended by gentle pipetting. Lysates were clarified by centrifugation at 13,000 rpm for 10 minutes at 4 °C before separation on a 10% SDS-PAGE gel.

For blotting analysis, gels were transferred to nitrocellulose membrane, stained by Ponceau S (10 minutes in 0.1% w/v Ponceau S in 5% acetic acid/water), and blocked with "blot blocking buffer" (3% w/v BSA and 0.1% Tween-20 in Tris-buffered saline) at 4 °C overnight. The blots were immersed in 0.3 μg/mL streptavidin-HRP (Thermo Scientific) at room temperature for 60 minutes, then rinsed with blot blocking buffer 4 × 5 minutes before development with SuperSignal West Pico reagent (Thermo Scientific) and imaging on an Alpha Innotech gel imaging system.

Electron microscopy of APEX fusions (related to Figs. 1E, 3, and S11)

HEK293T or COS7 cells were transfected with the APEX fusion construct of interest using Lipofectamine as described above. Fixation was performed with 2% glutaraldehyde in PBS as described previously (*5*). A solution of 3,3'-diaminobenzidine (DAB) was overlaid, APEX-catalyzed polymerization was allowed to proceed for 5 - 25 minutes, and DAB polymers were subsequently stained with OsO_4 as described previously (*5*). Cells were then rinsed 5 × 2 minutes in chilled distilled water and placed in chilled 2% aqueous uranyl acetate (Electron Microscopy Sciences) for 1 hour. Cells were brought to room temperature, washed in distilled water, then carefully scraped off the plastic, resuspended, and centrifuged at 700 g for 1 minute to generate a cell pellet. The supernatant was removed, and the pellet was dehydrated in a graded ethanol series (70%, 95%, 100%, 100%), for 10 minutes each time, then infiltrated in Eponate resin (Ted Pella) or EMBED-812 (Electron Microscopy Sciences) using 1:1 (v/v) resin and anhydrous ethanol for 1 hour, then 2:1 resin:ethanol overnight, then 100% resin for 2 hours. Finally, the sample was transferred to fresh resin and polymerized at 60 °C for 48 hours. Embedded cell pellets were cut with a diamond knife into 400 nm sections and imaged on a FEI-Tecnai™ G² Spirit BioTWIN transmission electron microscope operated at 80 kV.

For the model presented in Fig. 3E, human PPOX (residues 1-477) and human CPOX (residues 119-453) structures from PDB IDs 3NKS (*26*) and 2AEX (*28*) were used. For the model in Fig. 3F, tobacco PPOX and human FECH structures from PDB IDs 1SEZ (*23*) and 2QD1 (*29*) were used.

Two-color STORM super-resolution imaging (Fig. 1C)

U2OS cells were plated on Lab-Tek™ II Chambered Coverglass (Thermo Scientific) pre-coated with 10 μg/mL human fibronectin for 30 minutes. At 60-70% confluence, cells were transfected with mito-APEX, labeled with biotin-phenol and H_2O_2, and fixed and permeabilized as described above under "Characterization of APEX-mediated biotinylation by fluorescence imaging."

For two-color STORM, mito-APEX was detected by staining with mouse anti-V5 antibody (Invitrogen; 1:2000 dilution) in blocking buffer for 1 hour at room temperature. Cells were then washed 4 × 5 minutes with PBS and incubated with donkey anti-mouse-AF405/AF647 conjugate (1:2000 dilution). To detect biotinylated proteins, streptavidin-Cy3/Cy5 conjugate was added together with the secondary antibody, and the sample was incubated in blocking buffer for 5 minutes at room temperature. The antibody conjugate was prepared by coupling donkey anti-mouse antibody (Jackson Immunoresearch) with AF405-NHS and AF647-NHS esters (Invitrogen). The streptavidin conjugate was prepared by coupling homemade streptavidin protein (*27*) with Cy3-NHS and Cy5-NHS esters (Invitrogen). Labeled cells were washed 4 × 5 minutes with PBS.

Stained samples were imaged in STORM imaging buffer (10% glucose (w/v), 50 mM Tris pH 8.2, 10 mM NaCl, 0.56 mg/mL glucose oxidase, 0.8 mg/mL catalase, and 7.7 mg/mL cysteamine). STORM images were acquired on a home-built setup constructed on an Olympus IX-81 stand. A polychroic mirror (Di01-R405/488/561/635, Semrock) was used to reflect lasers onto the sample. A 642 nm laser line operating at maximum power (140 mW) was used to image AF647 and Cy5; 405 nm and 561 nm laser lines were used alternately to activate fluorophore pairs, and their intensities (from 0.1 mW to ~10 mW) were adjusted during the course of imaging experiment to maintain a steady level of fluorophore activation. Samples were illuminated by lasers in the pseudo-TIRF mode, and images were collected on an EMCCD camera (Evolve, Photometrics). Super-resolution images were typically reconstructed from more than 10,000 frames of localizations for each channel. Data were analyzed using Insight3 software from Xiaowei Zhuang (Harvard University).

Due to the use of overlapping reporter dyes (AF647 and Cy5) on the antibody and the streptavidin, we performed controls to check the degree of cross-talk between the two STORM channels. Control samples were prepared with only antibody labeling, or only streptavidin labeling, and imaged under identical conditions. By measuring fluorophore localization numbers we found that AF405/AF647 is activated by the 561 nm laser by 5.2% relative to activation by the 405 nm laser. Conversely, Cy3/Cy5 is activated by the 405 nm laser by 14.6% relative to activation by the 561 nm laser. These values represent the cross-talk between streptavidin and anti-V5 channels in Fig. 1C.

Proteomic mapping of the mitochondrial matrix

We followed the experimental configuration shown in Fig. S8A. First, to metabolically label the proteomes of HEK293T cultures with heavy or light isotopes of lysine and arginine (*18*), we split early passage HEK293T cells on day 0 into two T25 flasks at 10–15% confluence. One flask was cultured in light SILAC media: Dulbecco's Modified Eagle's Medium (DMEM) deficient in L-arginine and L-lysine (custom preparation from Caisson Laboratories) that we supplemented with L-arginine (Arg0) and L-lysine (Lys0) (Sigma-Aldrich) at concentrations of 84 mg/L and 146 mg/L, respectively, 10% dialyzed fetal bovine serum (Sigma-Aldrich), penicillin, streptomycin, glutamine (Gibco), and 4.5 g/L glucose. The second flask was cultured in heavy SILAC media with the same composition as above except Arg0 and Lys0 were replaced by L-arginine-$^{13}C_6$,$^{15}N_4$ (Arg10) and L-lysine-$^{13}C_6$,$^{15}N_2$ (Lys8) (Sigma Isotech). Every two days, before cells reached confluency, the heavy and light SILAC cultures were split into fresh SILAC heavy and light media, respectively. After three passages in T25 flasks, on day 6, light

and heavy SILAC cultures were expanded into T75 flasks at 20% confluence and cultured for two more days. We prepared two T75 flasks (H1, H2) for heavy SILAC cultures and three T75 flasks (L1, L2, L3) for light SILAC cultures.

On day 9, the H1, H2, L2, and L3 cultures containing 7-8 million cells each were transfected with 15 μg mito-APEX plasmid using 75 μL Lipofectamine in 30 mL light or heavy DMEM without serum or antibiotics. 4 hours after transfection, the media was replaced with fresh light or heavy SILAC media, and the cells were incubated overnight. On day 10, 18 hours after transfection, 500 μM biotin-phenol in 20 mL SILAC media was added to the H1, H2, L1, and L3 cultures. Cells were incubated for 30 minutes at 37 °C before addition of H_2O_2 (final concentration C_f 1 mM) for 1 minute at room temperature. Cells were then immediately quenched by aspirating off the biotin-phenol/H_2O_2 solution and replacing with a "quencher solution" consisting of 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in 20 mL DPBS. Cells were washed twice with the quencher solution and twice with DPBS to remove excess biotin-phenol. For the L2 culture (negative control with biotin-phenol and H_2O_2 omitted), cells were washed three times with DPBS. After the final wash, 7 mL of quencher solution was added to each flask, and cells were collected by gentle pipetting followed by centrifugation at 200 rpm for 5 minutes at room temperature. Supernatant was discarded, and the cell pellet was stored at -80 °C.

Cell pellets were lysed by thawing on ice, then adding 300 μL of freshly-prepared RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1× protease cocktail (Sigma Aldrich, catalog no. P8849), 1 mM PMSF, 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox) and incubating for 5 minutes at 4 °C. Lysates were clarified by centrifugation at 13,000 rpm for 10 minutes at 4 °C. Protein concentration in the supernatant was measured using the Pierce 660 nm Protein Assay kit, with bovine serum albumin as the reference standard.

Optimization of proteomic labeling conditions

The optimized proteomic labeling conditions described above were determined after performing the following tests. To optimize the H_2O_2 concentration, we first decided to restrict the H_2O_2 exposure time to 1 minute, to minimize cellular toxicity and maximize the temporal resolution of our method. Given this temporal restriction, we found that lowering the H_2O_2 concentration to <1 mM significantly reduced the biotinylation signal, assessed by imaging. We suspect this is the case despite the low K_m of APX for H_2O_2 (~20 μM (*30*)) and the high membrane permeability of H_2O_2 because endogenous catalase activity in mammalian cells rapidly consumes exogenously added H_2O_2.

To optimize biotin-phenol delivery, we tested loading for 1 minute, 3 minutes, 10 minutes, and 30 minutes and found much stronger labeling with the 30 minute pre-incubation prior to H_2O_2 addition. This may be because the biotin-phenol probe has limited membrane permeability. We also performed an empirical titration and found that concentrations below 500 μM significantly reduced biotinylation signal, assessed by imaging.

Finally, we found that it was critical to terminate the labeling after 1 minute with a quencher solution containing radical quenchers and peroxidase inhibitor. With ordinary cell lysis buffer, we observed biotinylation of outer mitochondrial membrane proteins such as TOM20 by matrix-localized APEX, probably because peroxidase-mediated

labeling continues even after dissolution of mitochondrial membranes. Upon addition of quenchers to the wash and lysis buffers, labeling of TOM20 and other contaminants was eliminated.

Enrichment of biotinylated proteins with streptavidin beads

The supernatants from lysed SILAC cultures above were combined as shown in Fig. S8A. H1 and L1 were mixed in a 1:1 ratio (by protein concentration) to give the Replicate 1 sample, and H2 and L2 were mixed in a 1:1 ratio to give the Replicate 2 sample. In addition, L3 supernatant was spiked in at 5% (by protein concentration) to each sample to facilitate quantitation of isotope ratios (to avoid infinite H/L values due to absence of light peptides).

Streptavidin-coated magnetic beads (Pierce catalog no. 88817) were prepared by washing twice with RIPA lysis buffer (formulation given above). 4 mg of total protein (at 4-5 mg/mL protein concentration) from each replicate sample was mixed with 500 µL of streptavidin bead slurry. The suspensions were incubated at room temperature for 1 hour with gentle rotation. Streptavidin beads were then washed with 2 × 1 mL RIPA lysis buffer, once with 1 mL of 2M urea in 10 mM Tris-HCl pH 8.0, and again with 2 × 1 mL RIPA lysis buffer. Biotinylated proteins were eluted by incubating the beads with 60 µL 1x NuPAGE LDS Sample Buffer (Invitrogen) supplemented with 20 mM DTT and 2 mM biotin and heating to 95 °C for 5 minutes.

In-gel protein digestion and extraction

Biotinylated proteins eluted from magnetic streptavidin beads were separated on a NuPAGE Novex Bis-Tris 4-12% gel run for 1 hour at 130V. The gel was stained overnight with Coomassie G-250 (Invitrogen) and subsequently destained with water. Lanes for each replicate were manually cut into 16 gel bands. Gel bands were destained in 1:1 acetonitrile:100 mM ammonium bicarbonate pH 8.0 for several hours followed by dehydration with acetonitrile. Dehydrated gel bands were swelled with 100 µL of 10 mM DTT in 100 mM ammonium bicarbonate for one hour while shaking. The DTT solution was subsequently removed and 100 µL of 55 mM iodoacetamide in 100 mM ammonium bicarbonate was added to each gel band and allowed to react for 45 minutes in the dark. The iodoacetamide solution was removed and bands were dehydrated with acetonitrile. Approximately 10-50 µL of 20 ng/µL Sequencing Grade Trypsin (Promega) was added to each sample and digestion was completed overnight with shaking at room temperature.

After the overnight digestion, excess trypsin solution was removed from each sample and bands were incubated with 20 µL of extraction solution (60% acetonitrile/0.1% TFA). The extraction solution was collected after 10 minutes and this process was repeated two additional times. The final peptide extraction was completed by incubating gel bands with 100% acetonitrile for 1-2 minutes. Extracted peptides were dried to completeness in a vacuum concentrator, then reconstituted in 100 µL of 0.1% formic acid and loaded onto C18 StageTips (*31*). The tips were washed twice with 50 µL of 0.1% formic acid, and peptides were eluted with 50% acetonitrile/0.1% formic acid, before drying in a vacuum concentrator.

Liquid chromatography-mass spectrometry

Dried peptides were reconstituted in 0.1% formic acid/3% acetonitrile and analyzed on a Q Exactive mass spectrometer (Thermo Scientific) coupled online to an Easy-nLC 1000 UPLC (Proxeon). Online chromatographic separation was completed using a capillary column (360 µm o.d. × 75 µm i.d.) with a 10 µm integrated emitter tip and self-packed with 23 cm of 1.9 µm ReproSil-Pur C18-AQ resin (Dr. Maisch GmbH). The UPLC mobile phase A was 0.1% formic acid/3% acetonitrile and the mobile phase B was 0.1% formic acid/90% acetonitrile. A linear gradient flowing at 200 nL/min and starting at 6% B and increasing to 30% B over 82 minutes was used to separate peptides with a total run time of 120 minutes.

The Q Exactive mass spectrometer was operated in the data-dependent mode such that after each MS1 scan, consecutive high energy collision dissociation MS2 scans were recorded on the 12 most abundant precursors. The resolution used to acquire MS1 and MS2 scans was 70,000 and 17,500, respectively. The MS1 advance gain control (AGC) target was set to 3×10^6 ions and the maximum ion time was set to 10 msec. MS2 scans were recorded with an AGC target of 5×10^4 ions with a maximum ion time of 120 msec, isolation width of 2.5 m/z, normalized collision energy of 25, dynamic exclusion time of 20 sec, underfill ratio of 5.0%, and peptide match and isotope exclusion enabled.

Proteins and peptides were identified and quantified using the MaxQuant software package (version 1.2.2.5) (*32, 33*). Data were searched against the UniProt human database which contains 81,470 entries. A list of 248 common laboratory contaminants as provided by the MaxQuant framework was also added to the database. A fixed modification of carbamidomethylation of cysteine and variable modifications of *N*-terminal protein acetylation, oxidation of methionine, and biotin-phenol on tyrosine were searched. The enzyme specificity was set to trypsin allowing cleavages *N*-terminal to proline and a maximum of two missed cleavages was utilized for searching. The maximum precursor ion charge state was set to 6. The precursor mass tolerance was set to 20 ppm for the first search where initial mass recalibration is completed and 6 ppm for the main search. The MS/MS tolerance was set to 50 ppm and the top MS/MS peaks per 100 Da was set to 10. The peptide and protein false discovery rates were set to 0.01 and the minimum peptide length was set to 6. Unmodified, *N*-terminally acetylated peptides, and peptides with oxidized methionines were used for protein quantification.

Cut-off analysis to determine the mitochondrial matrix proteome (related to Fig. S8)

Protein information was obtained from the MaxQuant Protein Groups table and is presented in Tables S1-S3. Proteins identified as reverse hits or contaminants were removed from the dataset. For each replicate, only proteins identified by three or more unique peptides with quantified SILAC ratios were retained for further analysis. Protein SILAC ratios were calculated by MaxQuant as the median of all corresponding peptide SILAC ratios (*32, 33*). Protein SILAC ratios from a given replicate were normalized such that the median H/L value was 1.00.

Proteins identified in each replicate experiment were sorted into one of three classes: (1) “mito” for those with previous mitochondrial annotation in MitoCarta (*16*), the Gene Ontology Cell Component database, or literature; (2) “false-positive” for those found in a hand-curated list of 2410 non-mitochondrial proteins (*16*); and (3) neither. Separate histograms were prepared for class (1) and (2) proteins, binned by $\log_2$(H/L)

SILAC ratio, as in Fig. S8C. We calculated the false positive rate (FPR) as a function of protein SILAC ratio using the equation:

$$FPR\ (\text{SILAC ratio}) = \frac{\text{P(SILAC ratio|false positive)}}{\text{P(SILAC ratio|mito)}}$$

The numerator is the conditional probability of finding a false positive protein in a particular SILAC ratio range and is calculated as the number of false positive proteins detected in that SILAC ratio range divided by the total number of false positive proteins detected in the entire dataset. Likewise, the denominator is the conditional probability of finding a known mitochondrial protein in a particular SILAC ratio range and is calculated similarly. The FPR is the ratio of the two conditional probabilities, and is plotted in Fig. S8D for Replicate 1.

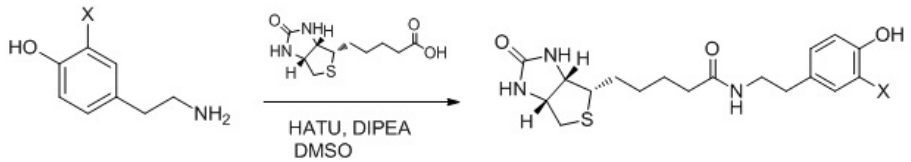
We selected as our cut-off point for each replicate the SILAC ratio at which the FPR reaches 0.1. All proteins above this cut-off are >10 times more likely to be mitochondrial than non-mitochondrial. The cut-offs (in $\log_2$(H/L) units) for Replicates 1 and 2 were 1.05 and 1.25, respectively. This gave 527 enriched proteins for Replicate 1 and 579 enriched proteins for Replicate 2. We intersected these two lists to obtain the 495 proteins that we define as our mitochondrial matrix proteome.

Localization of biotin-phenol labeling sites

Biotin-phenol site information was obtained from the MaxQuant evidence table and is presented in Table S4. Only modified peptides with a score >50 were retained for further analysis. SILAC ratios of modified peptides were not used in this analysis. For a given biotin-phenol-modified peptide, there can be multiple non-distinct peptides detected (due to spread in liquid chromatography retention time and charge state); we present the highest-scoring version across all non-distinct species in Table S4.

Synthesis of biotin substrates (related to Fig. S2)

Biotin conjugate **5** (biotin-XX-tyramide) was purchased from Invitrogen. Biotin conjugate **8** (EZ-link photoactivatable biotin) was purchased from Pierce. The other biotin conjugates were synthesized according to this general scheme:



All starting materials were purchased from Sigma Aldrich or Alfa Aesar except 3-nitrotyramine (PIN Chemicals, UK). Biotin (100 mg) was dissolved in 2 mL DMSO (dimethyl sulfoxide) at room temperature. Into this solution was slowly added 1.1 equivalents of HATU (2-(7-aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) and 3.0 equivalents of DIPEA (*N*,*N*-diisopropylethylamine). The mixture was stirred for 10 minutes at room temperature. Then 1.0 equivalent of the primary amine was added slowly, and the reaction was incubated overnight at room temperature with stirring. Products were purified by semi-preparative HPLC (Varian

Prostar model 210) using a C18 reverse phase column (Agilent Dynamax 250 × 21.4 mm, microsorb 300-5 C18) with a 5% to 95% gradient of acetonitrile in water over 30 minutes. Biotin conjugates **1, 2, 3, 4,** and **6** eluted at 18-20 minutes. Overall yields were 40-60%.

Biotin conjugate **7** was synthesized by hydrolysis of biotin conjugate **6**. Ten equivalents of NaOH were added to 100 mg of **6** in 1 mL methanol, and the mixture was allowed to stir overnight at room temperature. Methanol was then removed *in vacuo* and **7** was purified by HPLC as described above.

MS characterization of purified probes on an Applied Biosystems Q-TRAP instrument showed:
biotin conjugate **1**: calculated for $C_{18}H_{26}N_3O_3S$ $[M+H]^+$: 364.1695; found: 364.3
biotin conjugate **2**: calculated for $C_{18}H_{25}N_4O_5S$ $[M+H]^+$: 409.1546; found: 409.3
biotin conjugate **3**: calculated for $C_{18}H_{26}N_3O_4S$ $[M+H]^+$: 380.1644; found: 380.4
biotin conjugate **4**: calculated for $C_{19}H_{28}N_3O_4S$ $[M+H]^+$: 394.1801; found: 394.4
biotin conjugate **6**: calculated for $C_{20}H_{28}N_3O_5S$ $[M+H]^+$: 422.1750; found: 422.2
biotin conjugate **7**: calculated for $C_{19}H_{26}N_3O_5S$ $[M+H]^+$: 408.1593; found: 408.4

^{1}H-NMR for biotin conjugate **1** (500 MHz, d_6-DMSO): δ 1.27 (2H, m), 1.42-1.61 (4H, m), 2.04 (2H, t, J = 7.5 Hz), 2.57 (3H, m), 2.79 (1H, dd, J_1 = 7.5 Hz, J_2= 5 Hz), 3.05 (1H, dd, J_1 = 6 Hz, J_2= 2 Hz), 3.18 (2H, dd, J_1= 6.5 Hz, J_2= 6.5 Hz), 4.12 (1H, m), 4.30 (1H, m), 6.46 (1H, s), 6.56 (1H, s), 6.67 (2H, d, J = 8 Hz), 6.96 (2H, d, J = 8 Hz), 7.85 (1H, t, J = 5.5 Hz). ^{13}C-NMR for biotin conjugate **1** (125 MHz, d_6-DMSO): δ 25.502, 28.175, 28.358, 34.557, 35.396, 40.062, 55.625, 59.456, 61.283, 115.283, 129.618, 140.697, 155.921, 163.161, 172.288, 192.483.

Imaging assay to determine the membrane-permeability of the phenoxyl radical (Fig. S6)
HEK293T cells were transfected with ^{W41A}APX targeted to either the cytosol (NES) or inner leaflet of the plasma membrane (CAAX). 24 hours after transfection, cells were incubated with 500 μM alkyne-phenol and 1 mM H_2O_2 for 1 minute at room temperature. After washing cells three times with DPBS, click chemistry was performed on the cell surface by incubating with 10 μM AF647-picolyl azide conjugate and a pre-formed mixture of 200 μM $CuSO_4$, 1 mM THPTA ligand, and 2.5 mM sodium ascorbate in DPBS for 10 minutes at room temperature (*34*). Cells were quickly washed with DPBS three times and immediately imaged live by confocal microscopy as described above.

For comparison, click chemistry was also performed on fixed and permeabilized cells, in order to visualize the total population of alkyne-phenol-labeled molecules. For this labeling, cells were first blocked after fixation using 1% bovine serum albumin + 1% casein in PBS for 30 minutes at room temperature. Then cells were treated with 2.5 μM AF647-picolyl azide conjugate and a pre-formed mixture of 1 mM $CuSO_4$, 100 μM TBTA ligand, and 2.5 mM sodium ascorbate in the same blocking buffer for 1 hour at room temperature.

Gel filtration chromatography (Fig. S7B)
APEX and wtAPX were expressed in BL21-DE3 *E. coli* and purified by nickel affinity chromatography as previously described (*5*). Protein concentrations were measured using the BCA assay (Pierce). We found that purified APEX had an Abs_{405}/Abs_{280} ratio > 2.0, indicating high heme occupancy (*5*). Purified wtAPX had a lower ratio than this, so we

reconstituted the enzyme with purified heme as previously described (*5*) to obtain a final Abs_{405}/Abs_{280} ratio of 2.3. Gel filtration chromatography was performed on a Waters HPLC system using a Superdex 75 10/300 column (GE Healthcare) as described previously (*5*).

In vitro analysis of biotin-phenol reaction with purified amino acids (Fig. S7C)
Reactions were assembled as follows: 500 µM amino acid (all 20 were purchased from Sigma), 500 µM biotin-phenol, 1 mM H_2O_2, and 10 nM HRP (Sigma catalog no. P6782) in PBS pH 7.4. HRP was added last to initiate the reaction. After 10 minutes at room temperature, reactions were quenched by addition of sodium azide (C_f 1 mM). Product mixtures were analyzed by LC-MS. The LC was a Shimadzu UFLC-XR equipped with a 4.6 × 50 mm C18 column (Shimadzu catalog no. 220-91394-00) running a gradient of 5% to 95% acetonitrile in 0.1% trifluoroacetic acid/water over 20 minutes. The MS was an Applied Biosystems Q-TRAP instrument.

Imaging analysis of mitochondrial orphans (Fig. S10)
Six mitochondrial orphan genes (TRMT61B, TRUB2, RPUSD3, PREPL, PNPO, TRMT11) were obtained with C-terminal V5 tags in the pLX304 mammalian expression vector from the Broad hORFeome library. COS7 cells were transfected with these expression plasmids using Lipofectamine 2000. 24 hours later, cells were fixed and stained with anti-V5 antibody as described above under "Characterization of APEX-mediated biotinylation by fluorescence imaging". To visualize mitochondria, cells were also stained with rabbit anti-Tom20 antibody (Santa Cruz Biotechnology; 1:400 dilution) in blocking buffer for 1 hour at room temperature, followed by secondary goat anti-rabbit-AF647 conjugate (Invitrogen; 1:1000 dilution) in blocking buffer for 1 hour at room temperature. Labeled cells were washed 4 × 5 minutes with PBS before imaging. Confocal microscopy was performed in the same way as described above.

To quantify the extent of mitochondrial localization for each of the orphan proteins, the mitochondrial regions were first identified by automated image segmentation of the anti-Tom20 (AF647) channel. These regions were examined visually to ensure good overlap with the mitochondrial marker. The mean fluorescence intensity of the orphan protein (AF568) channel was then calculated for both regions overlapping with mitochondria ("mito") and regions not overlapping with mitochondria ("non-mito"). After subtracting background, the intensity ratios of "mito" over "non-mito" were calculated. For each orphan protein, 3-5 fields of view each containing 1-3 cells were analyzed in this way, and the mito/non-mito ratios were averaged.

Generation of cells stably expressing PPOX-APEX (related Fig. S11B).
HEK293T cells were transfected at ~70% confluence with PPOX-APEX (in pLX304) and two lentiviral packaging plasmids (dR8.91 and pVSVG (*16*)) using Lipofectamine 2000. 48 hours after transfection, the cell culture media containing lentivirus was collected and filtered through a 0.45 µm syringe filter. To a fresh HEK293T culture at ~70% confluence with ~200,000 cells was added 500 µL of the filtered lentivirus-containing media. Two days after infection, cells were transferred to complete FBS media containing 2 µg/mL blasticidin for selection. Infected cells were maintained in 2 µg/mL blasticidin-containing media for 7 days (three passages), before replating for EM imaging.

Cloning and mutagenesis

A variety of strategies were used to clone the APEX fusion constructs. In some cases, we PCR-amplified the APEX gene, digested with restriction enzymes, and ligated directionally into similarly-cut vectors (e.g., pcDNA3, pDisplay, pEGFP, or pTRC). In other cases, we used PCR overlap extension to fuse or insert the APEX gene to the gene of interest. The resulting PCR product was digested with restriction enzymes and pasted into cut vectors. This strategy was used to clone $APEX_{120}$-CPOX, $APEX_{70}$-CPOX, $APEX_{205}$-PPOX and $APEX_{14}$-CHCHD3. Lastly, we used Gateway technology (Invitrogen) to clone some constructs, such as PPOX-APEX.

In some APEX fusion constructs, a short linker was added between APEX and the gene of interest. This was achieved by designing longer primers for PCR amplification of the APEX gene. In all cases, the CMV promoter was used to drive expression in mammalian cells. The pDisplay vector was used for constructs targeted to the secretory pathway (e.g., ER lumen and cell surface); pDisplay contains an Igk chain signal sequence and the transmembrane domain of the PDGF receptor (which is deleted for ER lumen-targeted constructs). Occasionally, the pEGFP vector was used because it originally contained the gene of interest (as in the case of actin-APEX). In all other cases, pcDNA3 was used as a generic vector for expression of proteins targeted to the mammalian cell cytosol or organelles other than the secretory pathway.

Point mutants were generated using standard QuikChange mutagenesis (Stratagene). Primers 30-40 base pairs in length were designed to have melting temperatures between 78-80 °C, with one or two mismatched base pairs.

We avoided using the HA epitope tag in this work because it is tyrosine-rich, and hence likely to be damaged by the peroxidase/biotin-phenol reaction.

Genetic constructs

The two tables below summarize the plasmids used for peroxidase expression in mammalian and bacterial cells. Human CPOX, PPOX, CHCHD3, Sam50, and COASY genes were obtained from the Broad hORFeome library. Vimentin-APEX has been described (*5*).

Mammalian expression plasmids				
Name	**Features**	**Promoter/ Vector**	**Variants**	**Details**
mito-APEX	*NotI*-mito-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pCDNA3		Soybean APEX has 3 mutations relative to wtAPX: K14D, W41F, and E112K. wtAPX gene from soybean was a gift from Emma Raven (University of Leicester). mito matrix targeting sequence: MLATRVFSLVGKRAISTSVCVRAH (*5*) V5: GKPIPNPLLGLDST
PPOX-APEX	*NotI*-PPOX-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pCDNA3		

PPOX-APEX for stable cells	*attB1*-PPOX-*BamHI*-V5-APEX-Stop-*attB2*	CMV/ pLX304		
$APEX_{205}$-PPOX	*NotI*-$PPOX_{(1-205)}$-*NheI*-V5-APEX-*AflII*-10aa linker-$PPOX_{(206-477)}$-Stop-*XhoI*	CMV/ pCDNA3		10aa linker: GGSGGSGGSR
CPOX-APEX	*HindIII*-CPOX-*NotI*-APEX-Flag-Stop-*AflII*	CMV/ pCDNA3		Flag: DYKDDDDK
$APEX_{120}$-CPOX	*NotI*-$CPOX_{(1-120)}$-*NheI*-V5-APEX-*AflII*-10aa linker-$CPOX_{(121-453)}$-Stop-*XhoI*	CMV/ pCDNA3		10aa linker: GGSGGSGGSR
$APEX_{70}$-CPOX	*NotI*-$CPOX_{(1-70)}$-*NheI*-V5-APEX-*AflII*-10aa linker-$CPOX_{(71-453)}$-Stop-*XhoI*	CMV/ pCDNA3		10aa linker: GGSGGSGGSR
APEX-NES	*NotI*-Flag-APEX-NES-Stop-*XhoI*	CMV/ pCDNA3	W41X	NES: LQLPPLERLTLD(*35*)
IMS-APEX	*NotI*-$LACTB_{(1-68)}$-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pCDNA3		$LACTB_{(1-68)}$: MYRLLSSVTARAAATAGPAWDGG RRGAHRRPGLPVLGLGWAGGLGL GLGLALGAKLVVGLRGAVPIQS (*36*). The LACTB gene was a gift from Timothy A. Brown and David A. Clayton (HHMI Janelia Farm)
APEX-NLS	*NotI*-Flag-APEX-*EcoRI*-NLS-Stop-*XhoI*	CMV/ pCDNA3		NLS: PKKKRKVDPKKKRKVDPKKKRKV (*37*)
APEX-TM	*EcoRI*-ss-HA-*ApaI*-APEX-*SacII*-myc-TM-Stop-*NotI*	CMV/ pDisplay		ss: Igk chain signal sequence in pDisplay (Invitrogen). HA: YPYDVPDYA myc: EQKLISEEDL TM: transmembrane domain of PDGF receptor from pDisplay (Invitrogen)
APEX-KDEL	*EcoRI*-ss-*BglII*-APEX-V5-KDEL-Stop-*NotI*	CMV/ pDisplay		KDEL: ER retention motif
$C1_{(1-29)}$-APEX	*HindIII*-$C1_{(1-29)}$-*NotI*-APEX-V5-Stop-*XhoI*	CMV/ pCDNA3		$C1_{(1-29)}$: MDPVVVLGLCLSCLLLLSLW KQSYGGGKL (endoplasmic reticulum membrane anchor (*38*))
Connexin43-APEX	*NotI*-Cx43-12aa linker-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pCDNA3		12aa linker: GSKGSGSTSGSG
Actin-APEX	*NheI*-Flag-APEX-4aa linker-*XhoI*-actin-Stop-*BamHI*	CMV/ pEGFP		APEX is derived from pea (*5*) rather than soybean 4aa linker: SGLR actin is human β-actin
^{W41F}APX-NES	*NotI*-Flag-^{W41F}APX-NES-Stop-*XhoI*	CMV/ pCDNA3	W41A	

^{W41A}APX-CAAX	*NotI*-Flag-^{W41A}APX-*EcoRI*-CAAX-Stop-*XhoI*	CMV/ pCDNA3		CAAX: RSKLNPPDESGPGCMSCKCVLS (*39*)
HRP-TM	*EcoRI*-ss-HA-*ApaI*-HRP-*SacII*-myc-TM-Stop-*NotI*	CMV/ pDisplay		The HRP gene was a gift from Frances Arnold (Caltech)
HRP-KDEL	*EcoRI*-ss-*BglII*-HRP-V5-KDEL-Stop-*NotI*	CMV/ pDisplay		
PNPT1-APEX	*NotI*-PNPT1-*BamHI*-V5-APEX-*XhoI*	CMV/ pCDNA3		The human PNPT1 gene was a gift from Tsutomu Suzuki (University of Tokyo)
$APEX_{14}$-CHCHD3	*NotI*-$CHCHD3_{(1-14)}$-*NheI*-V5-APEX-*AflII*-10aa linker-$CHCHD3_{(15-227)}$-Stop-*XhoI*	CMV/ pCDNA3		10aa linker: GGSGGSGGSR
COASY-APEX	*NotI*-COASY-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pCDNA3		
mito-pBirA	*NotI*-mito-*BamHI*-V5-pBirA-Stop-*XhoI*	CMV/ pCDNA3		pBirA: R118G mutant of BirA (promiscuous biotin ligase)
pBirA-KDEL	*EcoRI*-ss-*BglII*-pBirA-KDEL-Stop-*NotI*	CMV/ pDisplay		
pBirA-NLS	*NotI*-Flag-pBirA-*EcoRI*-NLS-Stop-*XhoI*	CMV/ pCDNA3		

Bacterial expression plasmids		
Name	**Features**	**Promoter/Vector**
His_6-APEX	*NcoI*-His_6-APEX-Stop-*XbaI*	lacO/pTRC99
His_6-wtAPX	*NcoI*-His_6-wtAPX-Stop-*XbaI*	lacO/pTRC99

Discussion of labeling radius

A future challenge lies in carefully defining the radius of peroxidase-catalyzed labeling when it is not membrane-enclosed. An imaging readout following live cell labeling can be misleading because biotinylated proteins can themselves diffuse during the 1 minute labeling window. Previous EM studies on fixed cells suggest that the labeling radius of the biotin phenoxyl radical can be as small as 20 nm (*8, 9*), and in the live cell context the radius may be even smaller due to the presence of endogenous radical quenchers such as glutathione. It should also be possible to modulate labeling radius using electron withdrawing substituents or by co-application of radical scavengers to cells.

Supplementary Figures

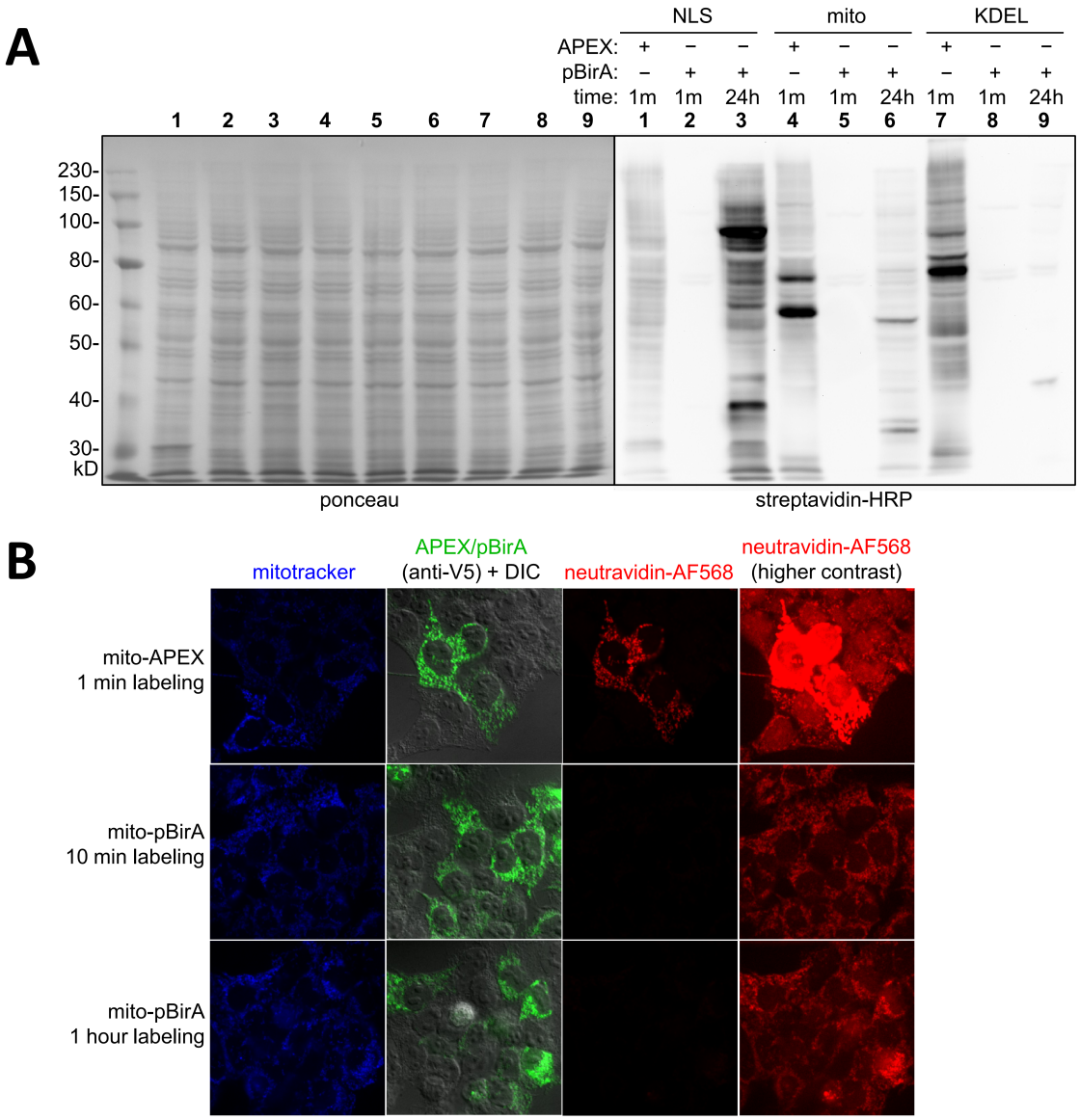


Fig. S1. Comparison of promiscuous biotinylation by APEX and pBirA. (**A**) Streptavidin blot analysis of whole cell lysate after live biotinylation by APEX or promiscuous biotin ligase (pBirA = R118G BirA (*3*)) in three different HEK293T cell compartments. The same samples were visualized by Ponceau stain on the left. APEX labeling was performed for 1 minute (1m) with 500 µM biotin-phenol and 1 mM H_2O_2. pBirA labeling was performed for 1 minute or 24 hours (24h) with 50 µM biotin, as previously described (*1*). NLS = nuclear localization signal. KDEL = endoplasmic reticulum retention sequence. mito = targeting sequence for the mitochondrial matrix. pBirA-catalyzed biotinylation is undetectable after 1 minute in the nucleus and mitochondrial matrix, and is very low in the lumen of the endoplasmic reticulum even after 24 hours, in contrast to APEX-catalyzed biotinylation. (**B**) Imaging analysis of biotinylation in the mitochondrial matrix. The top row shows 1 minute biotinylation by

mito-APEX, performed as in (A). In the bottom two rows, HEK293T cells were transfected with mitochondrial matrix-localized pBirA. 24 hours after transfection, 50 μM biotin was added to the culture medium to permit promiscuous biotinylation. After 10 minutes or 1 hour, cells were fixed and stained with neutravidin to detect biotinylated proteins and anti-V5 antibody (AF488 visualization) to detect pBirA expression. Mitotracker Deep Red staining was also used to visualize mitochondria. The neutravidin channel is shown at higher contrast in the rightmost column. The staining seen in rows 2 and 3 reflects endogenously biotinylated proteins (located in mitochondria (*40*)) because the staining is equivalent between transfected (V5-positive) and untransfected cells. Thus biotinylation of the mitochondrial matrix proteome by pBirA is insignificant after 1 hour, whereas APEX-catalyzed biotinylation in the same compartment is strong after 1 minute.

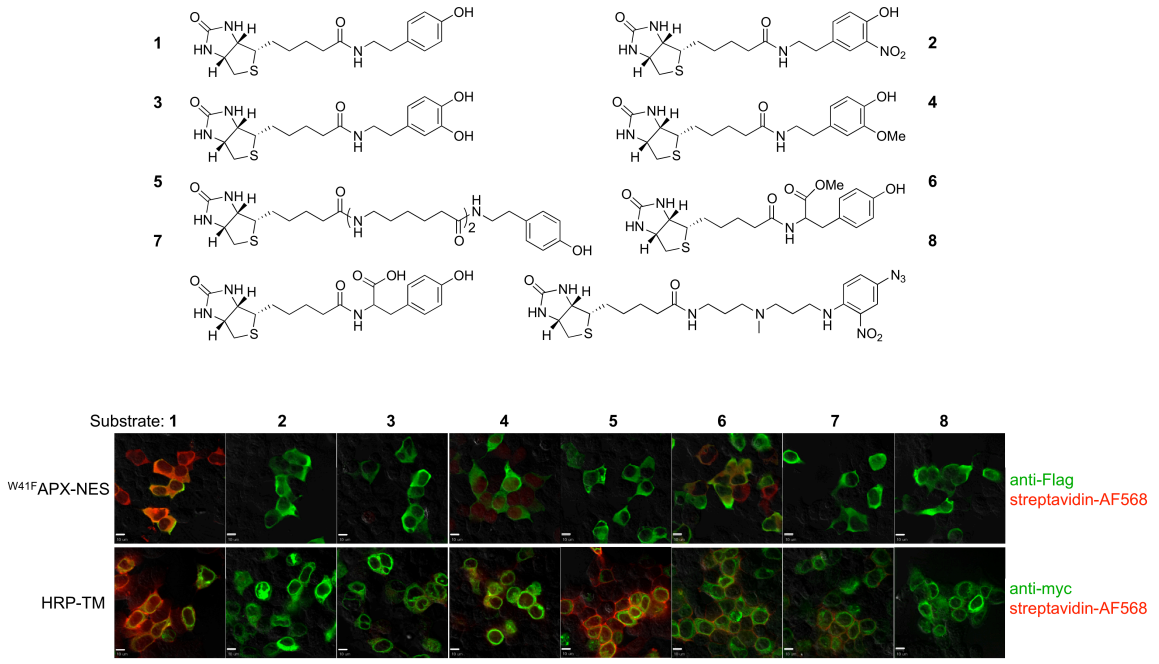


Fig. S2. Reactivity of eight different biotin conjugates towards APX and HRP. Top: Structures of biotin conjugates tested. These differ in electronic properties, linker length, sterics, and charge. Aryl azide **8** was used in a previous study with HRP and proposed to form a nitrene (*4*). Bottom: Imaging results with HEK293T cells expressing cytosolic Flag-^{W41F}APX-NES (top row) or cell surface HRP-myc-TM (bottom row). NES = nuclear export signal. TM = transmembrane domain of the PDGF receptor. For APX, labeling was performed by incubating cells with 500 μM of the indicated biotin substrate for 30 minutes, then adding 1 mM H_2O_2 for 1 minute to initiate biotinylation. Cells were then fixed and stained. For cell surface labeling with HRP, 100 μM of the indicated substrate was added for 10 minutes, then 1 mM H_2O_2 was added for 1 minute to initiate labeling. Cells were then fixed and stained. Anti-Flag/myc, streptavidin, and DIC images are overlaid. Labeling was not detected with HRP-TM and probe **8**, in contrast to a previous report (*4*). The simplest conjugate, biotin-phenol **1**, gave the strongest labeling of intracellular proteins, so we used this probe for all peroxidase-mediated biotinylation experiments in this work. Scale bars, 10 μm.

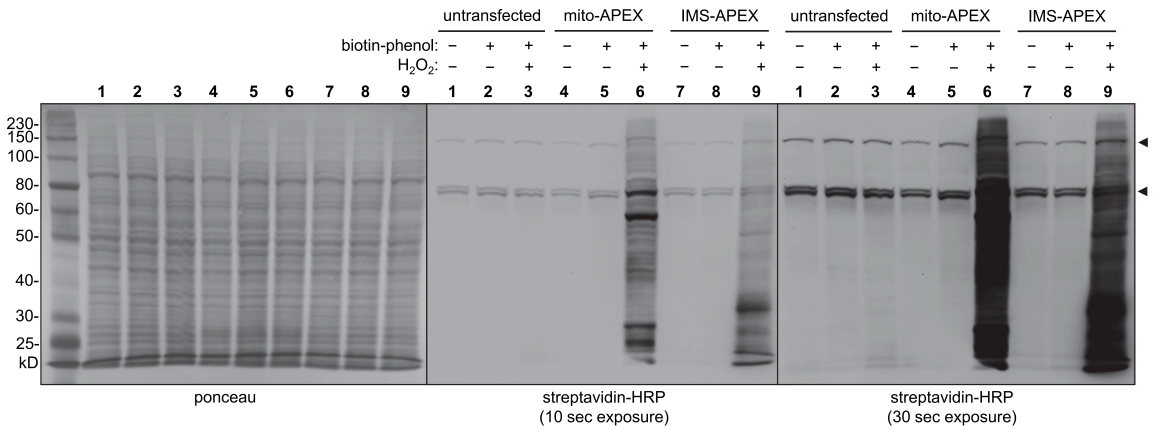


Fig. S3. APEX-catalyzed labeling requires addition of exogenous H_2O_2. HEK293T cells expressing APEX in the mitochondrial matrix (lanes 4-6) or the intermembrane space (IMS; lanes 7-9) were treated with biotin-phenol for 30 minutes and 1 mM H_2O_2 for 1 minute, as in the proteomic experiments, or with biotin-phenol alone for 30 minutes, or with neither reagent. Cells were then quenched and lysed, and whole cell lysate was analyzed by streptavidin-HRP blotting (two different exposures shown in middle and right blots). Labeling of many endogenous proteins was observed in lanes 6 and 9. No biotinylation above background was detected when biotin-phenol was added alone without exogenous H_2O_2 (lanes 5 and 8), indicating that endogenous H_2O_2 is not sufficient to start the reaction in either the matrix or the IMS. Untransfected HEK293T cells are included for comparison in lanes 1-3. Arrowheads point to endogenously biotinylated proteins (*40*).

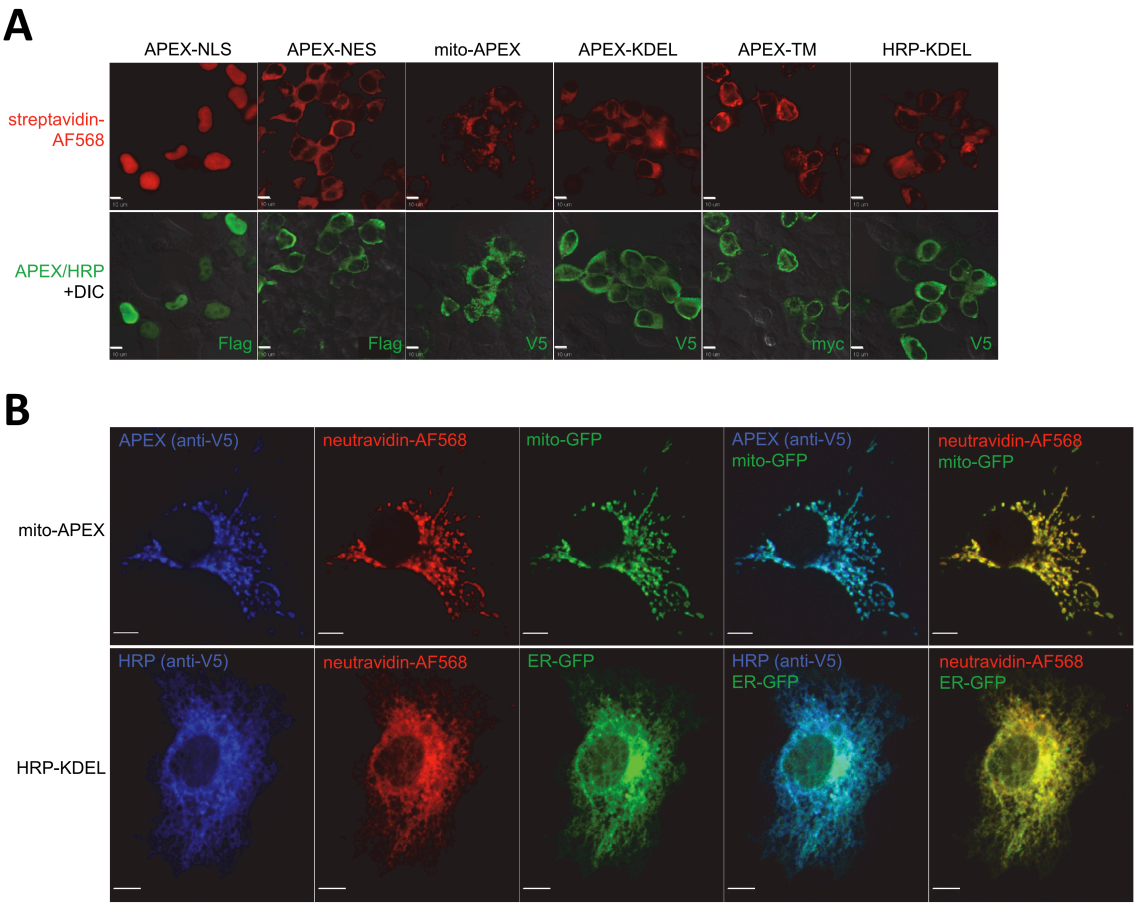


Fig. S4. Imaging analysis of APEX- and HRP-catalyzed biotinylation in multiple cellular compartments. (**A**) HEK293T cells were transfected with APEX or HRP constructs targeted to various compartments. Labeling was performed with biotin-phenol and H_2O_2 for 1 minute. Cells were then fixed and stained with streptavidin-AF568 to visualize biotinylated proteins, and antibody to visualize Flag, V5, or myc tags on the APEX/HRP enzymes. HRP-NES and mito-HRP constructs did not give detectable streptavidin staining, as expected (data not shown). NLS = nuclear localization signal; NES = nuclear export signal; mito = mitochondrial matrix signal; KDEL = endoplasmic reticulum retention motif; TM = transmembrane helix of PDGF receptor. Specific targeting sequences are given in the Materials & Methods. (**B**) Additional imaging analysis of endogenous proteins biotinylated by mito-APEX and HRP-KDEL, with respect to organelle markers. mito-GFP is GFP appended to a mitochondrial matrix signal. ER-GFP is GFP targeted to the endoplasmic reticulum membrane via fusion to the first 29 amino acids of cytochrome P450 (*38*) (this plasmid was a gift from Erik Snapp, Albert Einstein College of Medicine). COS7 cells were labeled and stained as in (A). Merged images are shown in the right two columns. Tight overlap is observed between proteins biotinylated by APEX/HRP and the respective organelle markers. All scale bars, 10 µm.

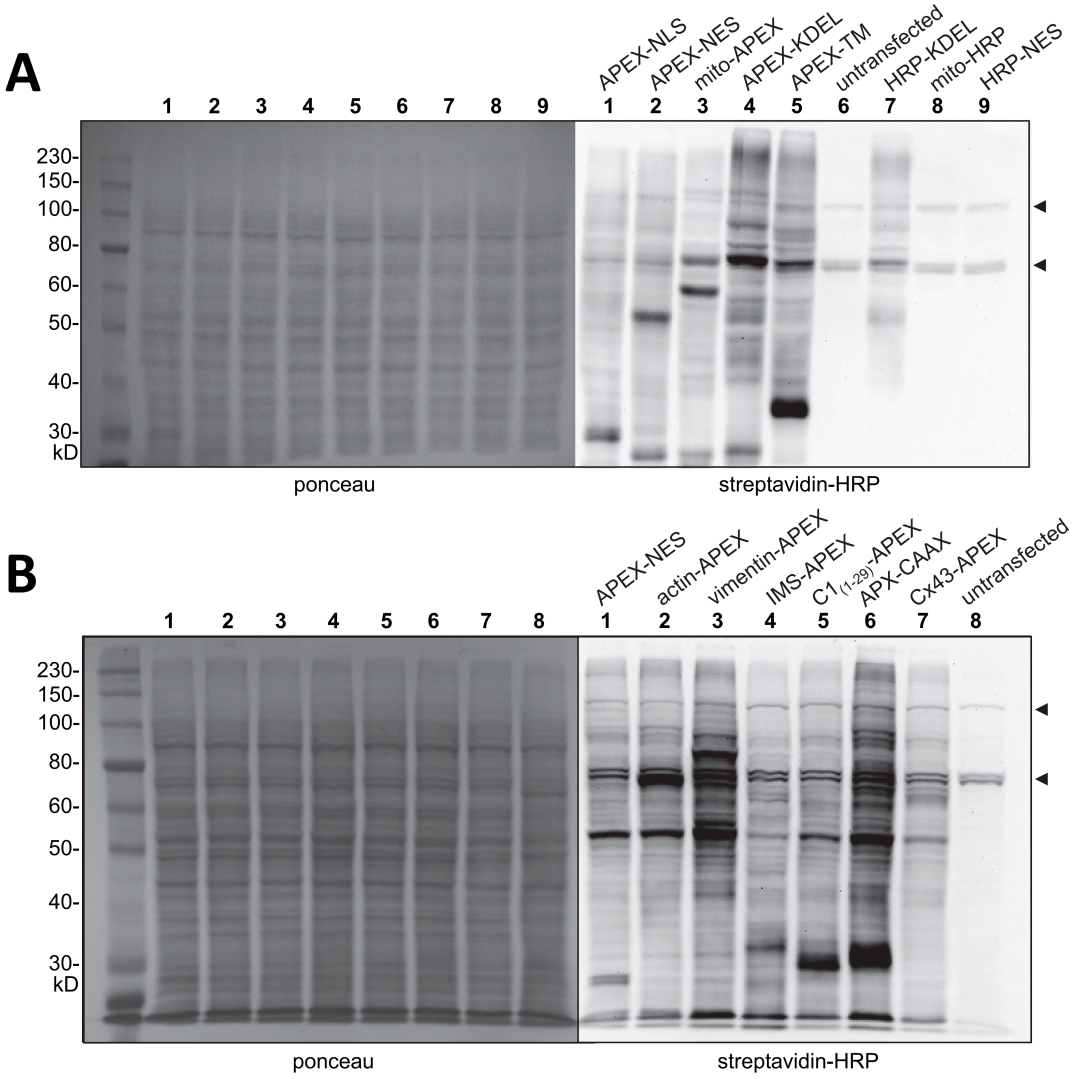


Fig. S5. Streptavidin blot analysis of APEX- and HRP-catalyzed biotinylation in many cellular compartments. (**A**) The same constructs as in Fig. S4A were introduced into HEK293T cells, which were labeled with biotin-phenol and H_2O_2 as in Fig. S4, then lysed. Whole cell lysates were analyzed by 10% SDS-PAGE, with Ponceau S staining (left) and streptavidin-HRP visualization of biotinylated proteins (right). Arrowheads point to endogenously biotinylated proteins (*40*). (**B**) Same analysis as in (A), for fusion constructs targeted to different regions of the cytosol. NES = nuclear export signal; IMS = mitochondrial intermembrane space signal; $C1_{(1\text{-}29)}$ = endoplasmic reticulum membrane anchor (APEX faces the cytosol) (*38*); CAAX = plasma membrane anchor (APEX faces the cytosol); Cx43 = connexin 43. Note that the IMS freely exchanges components <5 kD (including the biotin-phenoxyl radical) with the cytosol via porins in the outer mitochondrial membrane (*41*). Pea APEX (*5*) instead of soybean APEX was used for the vimentin and actin fusions. Soybean ^{W41A}APX was used for the CAAX fusion. For each of the organelle-targeted constructs in (A), and the cytosol-facing constructs in (B), a unique "fingerprint" of biotinylated endogenous proteins is observed, illustrating the spatial specificity of the labeling.

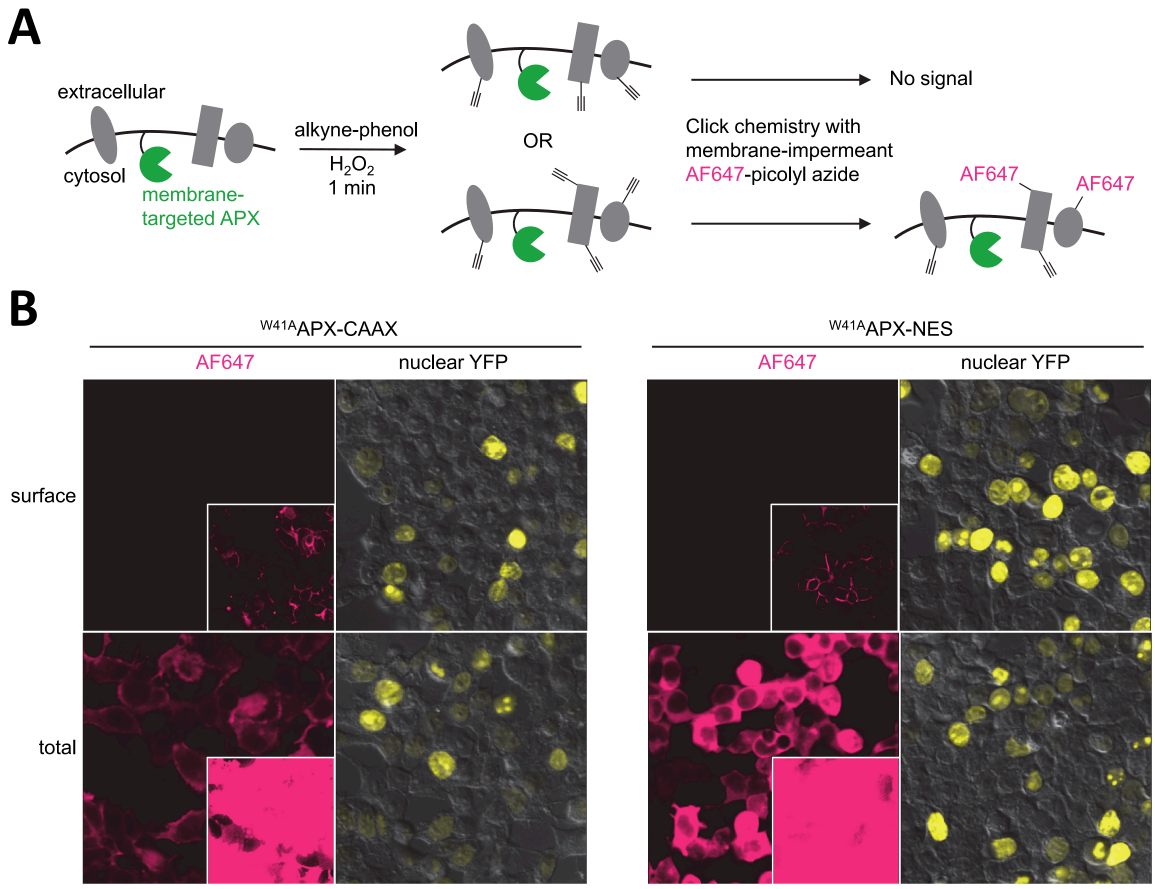


Fig. S6. Peroxidase-generated phenoxyl radicals do not cross the plasma membrane. (**A**) Assay scheme. HEK293T cells were transfected with ^{W41A}APX targeted to the cytosol (NES) or inner leaflet of the plasma membrane (CAAX; green pacman). Labeling was performed with alkyne-phenol and H_2O_2 for 1 minute. Thereafter click chemistry was performed at the cell surface with membrane-impermeant AF647-picolyl azide (*34*). Only if the phenoxyl radical crosses the plasma membrane will alkyne be present at the cell surface and detectable by the AF647 reagent. Dimeric ^{W41A}APX was used in this experiment because it has higher activity in the mammalian cytosol than monomeric APEX. (**B**) Images with ^{W41A}APX-CAAX (left) and ^{W41A}APX-NES (right). Nuclear-localized YFP was a transfection marker. As a control (bottom rows labeled "total"), cells were fixed and permeabilized prior to click chemistry to detect intracellular alkyne-phenol labeling. Insets show the same fields of view with 50-fold greater contrast. Quantitation shows that the signal at the cell surface (top rows) represents <2% of the total signal (bottom rows). The same results were obtained when we used biotin-phenol instead of alkyne-phenol and detected labeling at the cell surface with membrane-impermeant streptavidin-AF568 (data not shown). The alkyne-phenol experiment is shown here because it is more rigorous; detection of cell surface labeling by biotin-phenol, especially near the lipid bilayer, could be impaired by the large size of streptavidin.

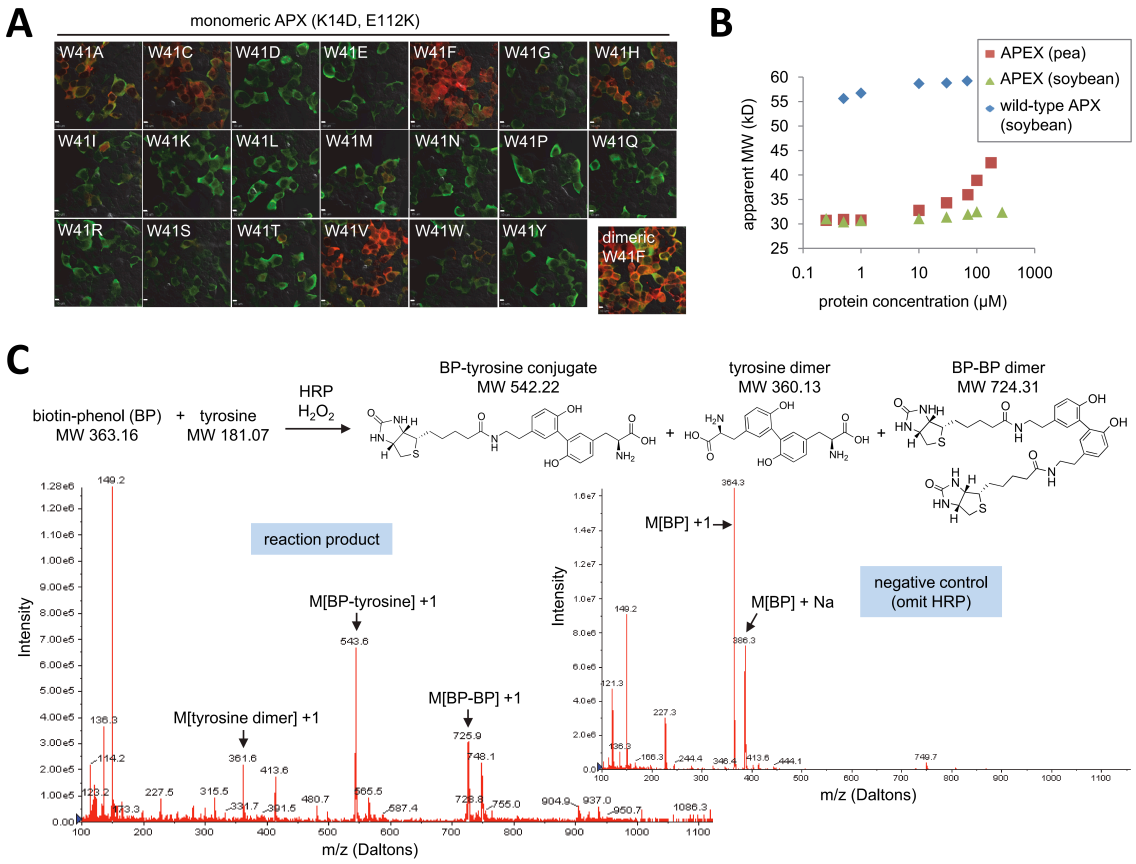
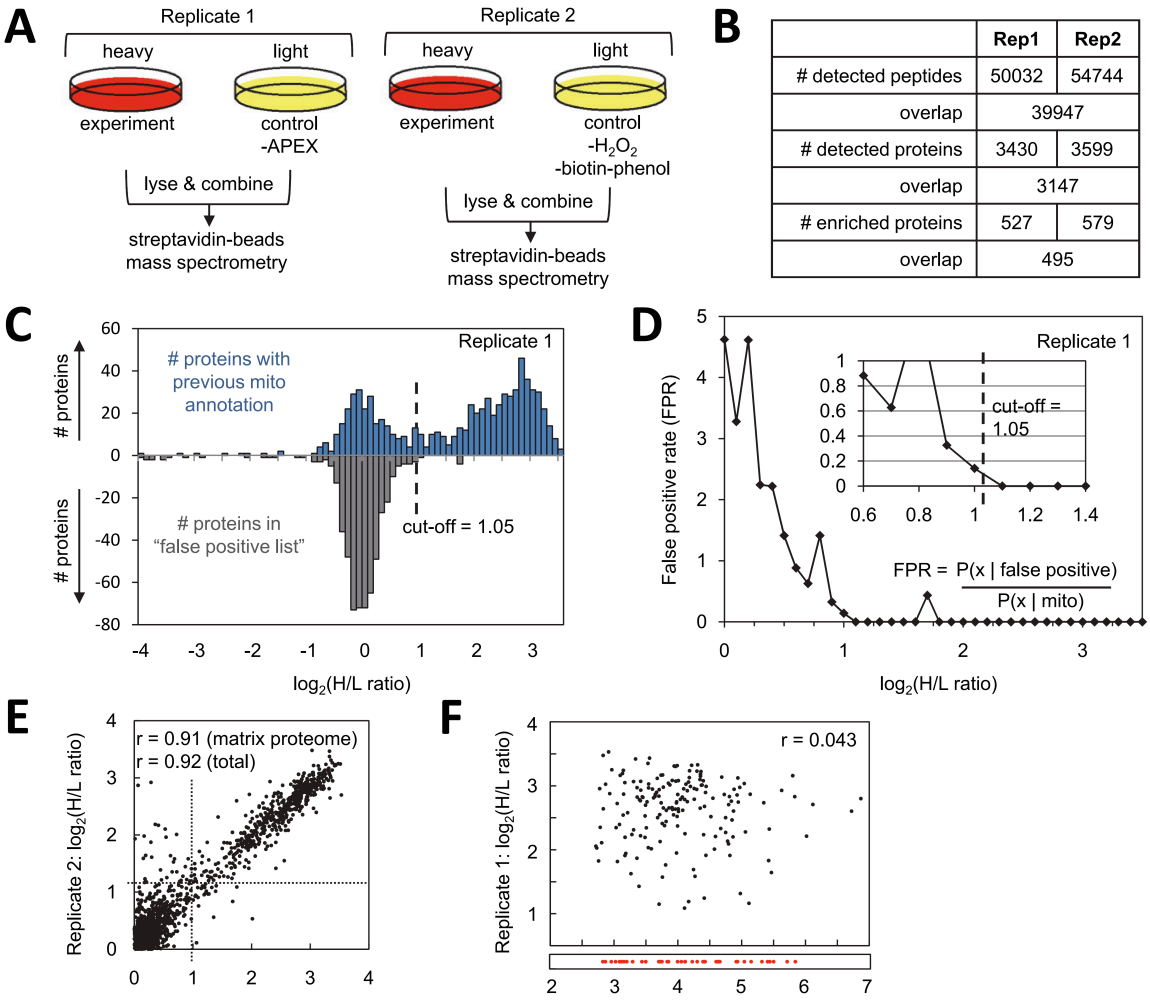


Fig. S7. Peroxidase engineering and characterization of monomericity and amino acid specificity *in vitro*. (**A**) The APEX construct previously introduced as a genetic tag for electron microscopy was derived from pea ascorbate peroxidase (APX) (*5*). This construct had three mutations relative to wild-type pea APX: two to make it monomeric (K14D, E112K), and one to boost activity towards aromatic substrates (W41F). In this work, we discovered that K14D and E112K had a more strongly monomerizing effect on APX from soybean than from pea (both wild-type enzymes are constitutive dimers (*42, 43*)). We therefore switched to the soybean enzyme. In addition, we performed comprehensive mutagenesis at the W41 active site residue to maximize intracellular activity towards biotin-phenol. Confocal images are shown for HEK293T cells expressing all possible W41x mutants (fused to nuclear export sequences, and on a monomeric soybean APX background). After labeling for 1 minute with biotin-phenol and H_2O_2, cells were fixed and stained with streptavidin-AF568 to detect biotinylated proteins (red) and anti-Flag to detect peroxidase expression (green). The same experiment with dimeric ^{W41F}APX (soybean) is shown for comparison. Anti-Flag, streptavidin, and DIC images are overlaid. From this experiment, we concluded that the W41F mutation was best, and combined it with the monomerizing mutations to give soybean APEX (K14D, W41F, E112K). Note that soybean-derived APEX was used for all EM and proteomic labeling experiments in this work, except where indicated otherwise. Scale bars, 10 μm. (**B**) Gel filtration chromatography was used to analyze the monomericity of soybean-derived APEX, pea-derived APEX from our previous study (*5*), and wild-type

soybean APX (calculated molecular weight 29 kD). As expected (*43*), wild-type APX runs as a dimer at all protein concentrations tested (250 nM–68 μM; apparent MW ~58 kD). Soybean APEX runs as a monomer (apparent MW ~30 kD), even at 270 μM, while pea APEX runs as a monomer at lower concentrations but shows signs of dimerization at higher concentrations, as previously observed (*5*). (**C**) HRP, biotin-phenol (BP), and H_2O_2 were combined in 20 separate *in vitro* reactions with each of the amino acids (at 500 μM). Reactions were quenched after 10 minutes at room temperature and analyzed by HPLC and MS. HRP was used because it has higher *in vitro* activity than APEX (*5*). In all reactions, BP dimer was detected. In the reaction with tyrosine, a BP-tyrosine adduct and tyrosine dimer with the masses and predicted structures shown were also detected. Negative controls with HRP or H_2O_2 omitted eliminated the formation of these products. In the reaction with cysteine, the amino acid was consumed but we did not detect the expected cysteine-BP adduct or cysteine dimer. No reaction was detected for the 18 other samples (apart from BP dimerization). We conclude that the reaction of BP with tyrosine is the easiest to detect (see also Fig. S9), but reactions at other electron-rich sidechains are also possible and have literature precedent (*10-12*). We note that based on analysis of available protein structures, >90% of proteins are expected to have one or more surface-exposed tyrosines[1]. Thus, even if phenoxyl radicals have a strong preference for labeling at tyrosine, this specificity should not limit by much the potential coverage of this technique.

[1]. This value was calculated by determining the number of Tyr residues for each protein in the human proteome, and applying the statistic that 80% of Tyr are surface-exposed (*14*). Hence, for the 677 (3.3%) proteins with 1 Tyr, 80% of them will have a surface Tyr. For the 813 (4%) proteins with 2 Tyr, 96% of them will have one or more surface Tyr, and so on.



	Rep1	Rep2
# detected peptides	50032	54744
overlap	39947	
# detected proteins	3430	3599
overlap	3147	
# enriched proteins	527	579
overlap	495	

Fig. S8. Determination of the mitochondrial matrix proteome. (**A**) Experimental setup. We established "heavy" and "light" isotopically labeled HEK293T cell cultures using the SILAC approach (*18*). Two replicate heavy SILAC cultures were labeled by pre-loading biotin-phenol for 30 minutes, then adding 1 mM H_2O_2 for 1 minute to initiate biotinylation. Two light SILAC cultures were processed in parallel as negative controls, one with APEX omitted, and one with biotin-phenol and H_2O_2 omitted. After lysis, heavy and light lysates were combined, enriched with streptavidin beads, eluted, separated by gel, digested to peptides with trypsin, and analyzed by tandem MS. (**B**) Table showing numbers of detected peptides, detected proteins, and enriched proteins for each replicate, as well as the overlap between the two replicates. We required three or more unique quantified peptides to consider a protein "detected". (**C**) Distribution of SILAC ratios ($\log_2$(H/L)) from the Replicate 1 experiment. The top panel shows "true positive" analysis: the number of proteins within each SILAC ratio range (bin size 0.1) with prior mitochondrial annotation. The bottom panel shows "false positive" analysis: the number of proteins within each SILAC ratio range appearing in a hand-curated list of 2410 non-mitochondrial proteins ("false positive list") (*16*). (**D**) False positive rate (FPR) as a function of SILAC ratio. At a given SILAC ratio, the FPR measures the probability of finding a false positive versus a true mitochondrial protein. We selected an FPR of 0.1 as

our cut-off point, i.e. 10 times more likely for a protein to be mitochondrial than a false positive. Replicate 1 had 527 proteins above this cut-off, and Replicate 2 had 579 proteins. The overlap between these groups was 495 proteins, which we defined as our matrix proteome (Table S1). (**E**) Correlation between SILAC ratios for Replicates 1 and 2. R values are shown for all detected proteins, and for proteins in our matrix proteome (proteins above and to the right of the dotted lines). (**F**) Correlation between SILAC ratio (from Replicate 1) and protein abundance in human U2OS cells (*44*), for proteins in our matrix proteome. The bottom strip, in red, shows protein copy number for matrix proteins from Fig. 2B that were *not* detected in our matrix proteome.

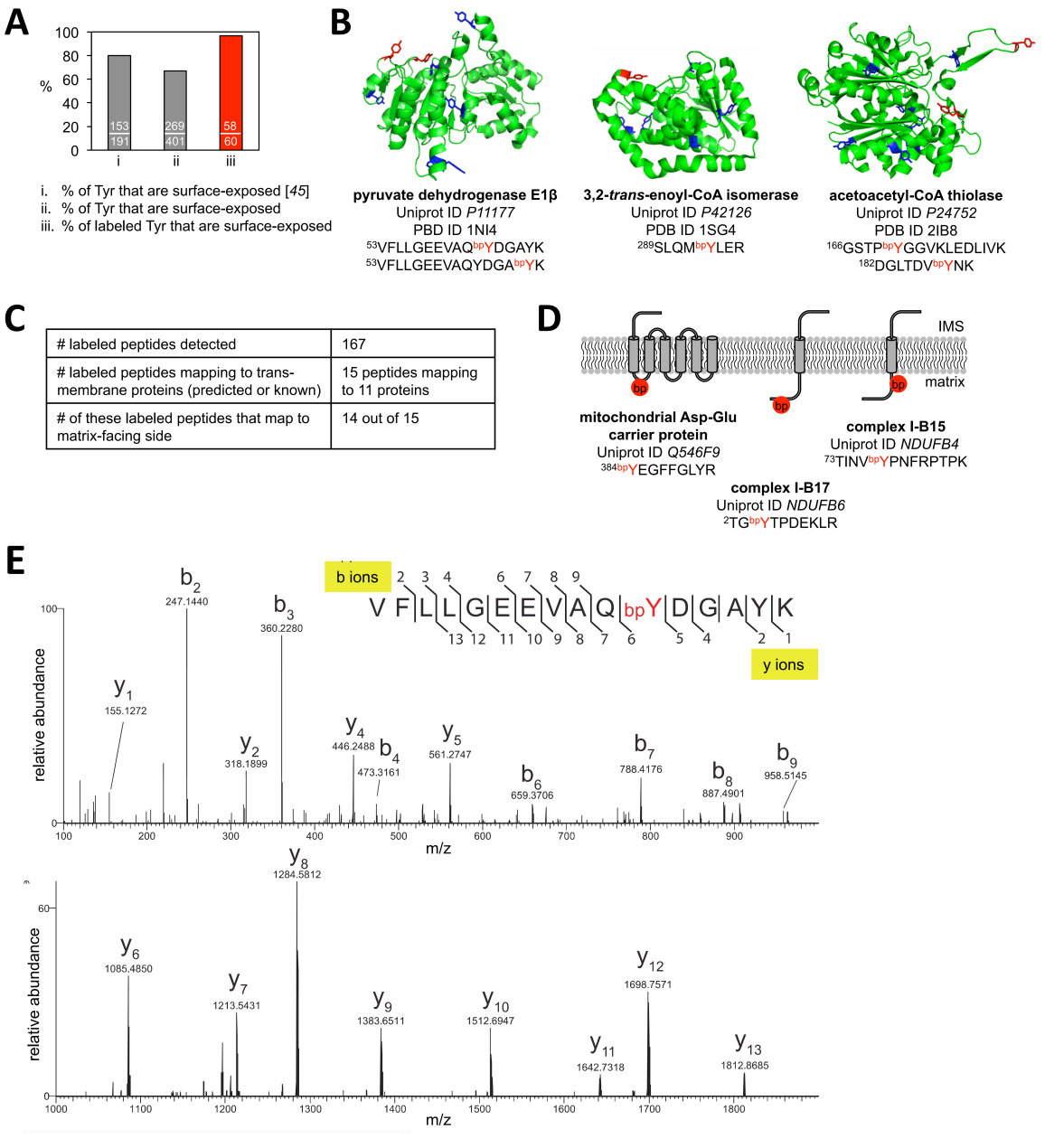


Fig. S9. Mapping biotinylated peptides to transmembrane and soluble matrix proteins. (**A**) Occurrence of surface-exposed tyrosines in soluble proteins with known structure. (i) is from reference (*45*). (iii) is based on analysis of the 60 biotinylated peptides we detect that map to soluble proteins with known structure (39 proteins). (ii) is based on all the tyrosines (biotinylated or not, 401 in total) within these 39 proteins. The criterion for "surface-exposed" is having >10 Å^2 solvent-accessible area, as calculated by PyMOL (*45*). This analysis shows that peroxidase-catalyzed biotinylation favors surface exposed tyrosine sites on soluble proteins. Protein structures were obtained by searching UniProt IDs in the Swiss Model Repository. Structures with >85% sequence homology to our detected proteins were used for this analysis. (**B**) Protein structures showing the positions of five of the biotinylated tyrosines (colored red, bp = biotin-phenol

modification) analyzed in (A). Non-labeled tyrosines within each protein are colored blue. Sequences of detected biotinylated peptides are shown beneath each protein structure. (**C**) Analysis of biotinylated peptides that map to transmembrane proteins with known or predicted topologies (according to the UniProt database). Nearly all of the biotinylation sites (14 out of 15) map to the matrix-exposed portion of these proteins. (**D**) Specific examples from (C). Biotinylation sites are depicted by the red circle. Sequences of detected biotinylated peptides are given underneath. (**E**) MS/MS spectrum recorded on the SILAC heavy-labeled $[M + 2H]^{2+}$ ion (m/z 1086.54) corresponding to the pyruvate dehydrogenase E1β peptide VFLLGEEVAQ^{bp}YDGAYK shown in (B). The measured precursor mass is within 1 p.p.m. of the calculated mass. Observed singly charged fragment ions are labeled in the spectrum and indicate that Y63 is modified with biotin-phenol. Details for all 167 biotinylated peptides are provided in Table S4.

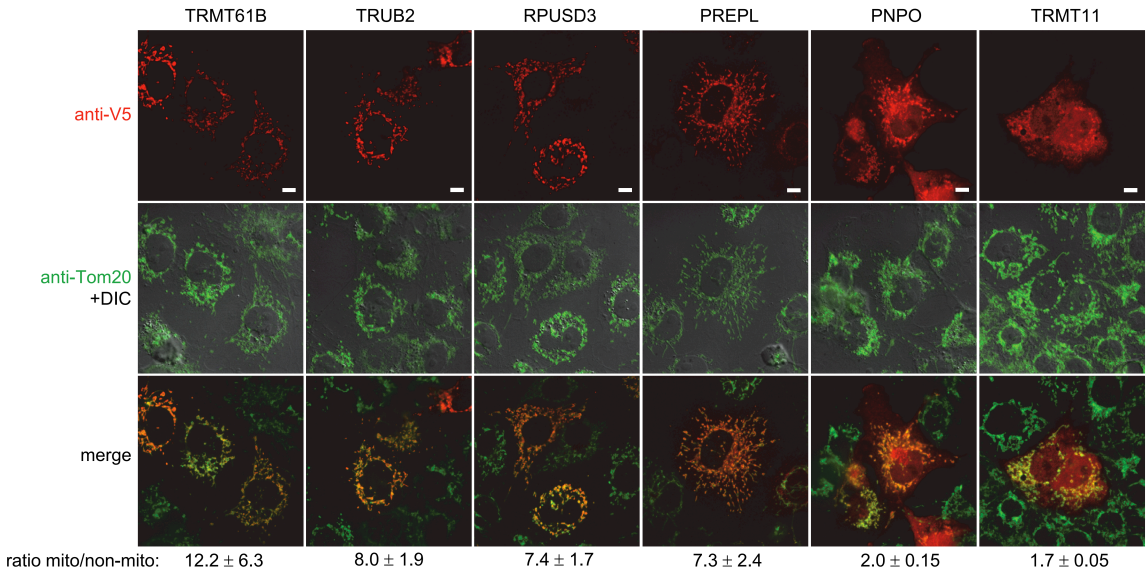


Fig. S10. Fluorescence imaging shows mitochondrial localization of six mitochondrial orphans. Six proteins selected at random from the 31 mitochondrial orphans (Table S2) identified in our matrix proteome were V5-tagged and expressed in COS7 cells. Protein localization was detected by anti-V5 staining (AF568). Mitochondria were visualized by anti-Tom20 staining (AF647). The ratio of V5 signal intensity overlapping with mitochondria versus not overlapping with mitochondria was quantified (bottom). For reference, the mito/non-mito ratio was 5.4±2.0 for mito-APEX (a positive control) and 1.4±0.16 for a cytosolic fluorescent protein (mCherry) fusion (a negative control). Four of these six orphan proteins are related to tRNA editing: TRMT61B and TRMT11 are tRNA methyltransferases, and TRUB2 and RPUSD3 are pseudouridylate synthases. TRUB2 has a yeast homolog, PUS4, with mitochondrial annotation (*46*). TRUB2 and RPUSD3 both have predicted mitochondrial targeting sequences. PREPL is prolyl endopeptidase-like protein. Though annotated as a cytosolic protein, a previous study found 35% of prolyl endopeptidase activity to be associated with crude mitochondria (*47*). PNPO is pyridoxamine-phosphate oxidase, which catalyzes the last and rate-determining step of vitamin B_6 synthesis, thought to occur in the cytosol. PNPO has a yeast homolog, PDX3, which was identified in an IMS proteomic study (*48*), but not validated biochemically. Scale bars, 15 μm.

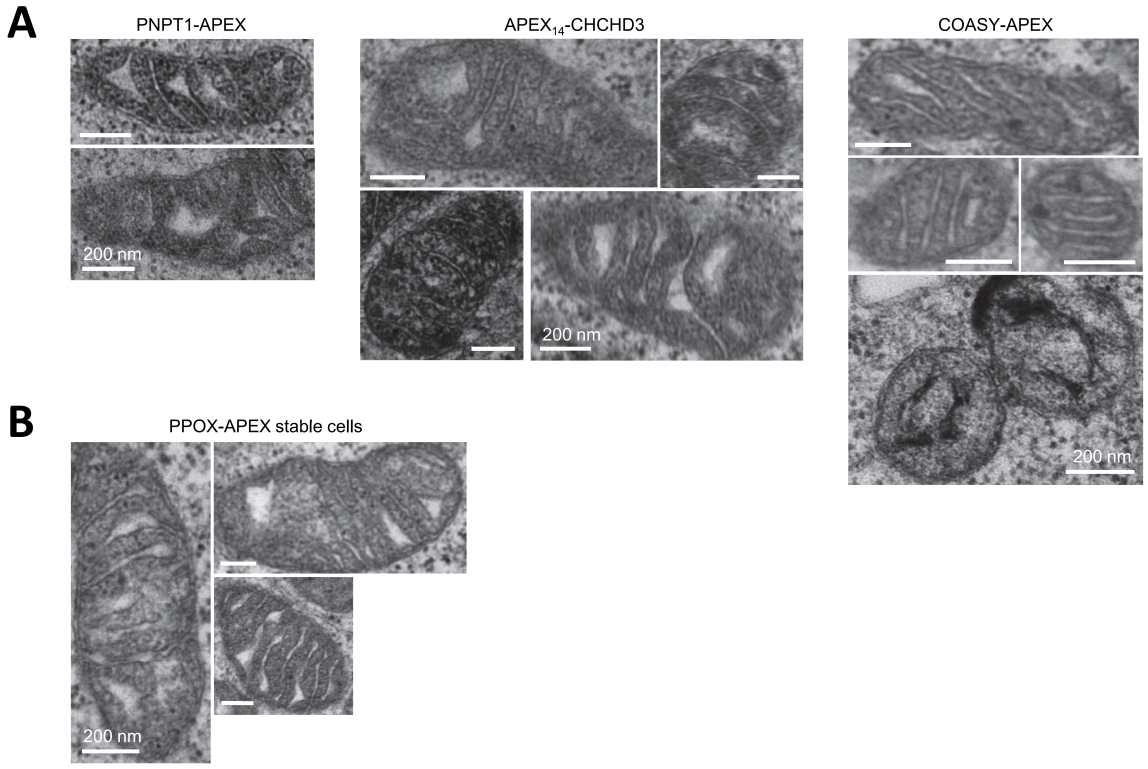


Fig. S11. Electron microscopy of APEX fusions to PNPT1, CHCHD3, COASY, and PPOX. (**A**) PNPT1 and CHCHD3 were previously localized to the mitochondrial intermembrane space (IMS) (*49, 50*), and COASY was previously localized to the outer mitochondrial membrane (OMM) (*51*), yet we detected all three proteins in our mitochondrial matrix proteome. APEX fusions to PNPT1 (C-terminal fusion) and CHCHD3 (insertion between amino acids 14 and 15) gave mitochondrial matrix staining by EM in HEK293T and COS7 cells, respectively. PNPT1 (polyribonucleotide nucleotidyltransferase 1) is an RNA binding and processing protein. In addition to our EM data which localizes the C-terminus to the matrix, we detected two PNPT1-derived biotinylated peptides (Table S4) that indicate that Tyr302 and Tyr356 are both matrix-exposed. CHCHD3 (coiled-coil-helix-coiled-coil-helix domain-containing protein 3) plays a role in bridging the outer and inner mitochondrial membranes (*52*), so perhaps it is not surprising that a portion of it (residue 14, according to our EM data) is matrix-exposed. A C-terminal APEX fusion to COASY (bifunctional coenzyme A synthase) in HEK293T cells gave matrix staining in some fields of view (top two images), and IMS staining in other fields of view (bottom). (**B**) Same experiment as in Fig. 3C, except that PPOX-APEX was stably rather than transiently expressed. These stable HEK293T cells were prepared by lentiviral infection of the PPOX-APEX gene and selection in 2 μg/mL blasticidin. All scale bars, 200 nm.

Supplementary Spreadsheet 1. Mitochondrial matrix proteome.
Table S1: Our mitochondrial matrix proteome (495 proteins), ranked from most enriched to least enriched, according to $\log_2$(H/L) SILAC ratio from replicate 1. 93 additional proteins below our cut-off point are also included but shaded grey.
Table S2: Mitochondrial orphans (31 proteins), ranked from most enriched to least enriched.
Table S3: Mitochondrial matrix orphans (240 proteins), ranked from most enriched to least enriched. These proteins have prior mitochondrial annotation but were not previously known to be localized to the matrix.
Table S4: Biotinylated peptides detected (167 peptides derived from 103 proteins). The peptides are grouped by protein and ranked by protein SILAC ratio.
Table S5: Column definitions.

Supplementary Spreadsheet 2. Analysis of protein groups and membrane complexes.
Table S6: Sub-mitochondrial localization of detected proteins. Related to Fig. 2A.
Table S7: Mitochondrial matrix protein groups detected. Related to Fig. 2B (analysis of depth of coverage). Eight proteins in these groups that were not detected by RNA-Seq in HEK293T cells (*53*) were not included in the analysis.
Table S8: Inner mitochondrial membrane protein complexes detected. Related to Figs. 2C-D.
Table S9: Column definitions.

References and Notes

27. M. Howarth, A. Y. Ting, Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc.* **3**, 534 (2008).
28. D. S. Lee *et al.*, Structural basis of hereditary coproporphyria. *Proc.Natl. Acad. Sci. U.S.A.* **102**, 14232 (2005).
29. A. E. Medlock, T. A. Dailey, T. A. Ross, H. A. Dailey, W. N. Lanzilotta, A pi-helix switch selective for porphyrin deprotonation and product release in human ferrochelatase. *J. Mol. Biol.* **373**, 1006 (2007).
30. D. A. Dalton, L. Diaz del Castillo, M. L. Kahn, S. L. Joyner, J. M. Chatfield, Heterologous expression and characterization of soybean cytosolic ascorbate peroxidase. *Arch. Biochem. Biophys.* **328**, 1 (1996).
31. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896 (2007).
32. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367 (2008).
33. J. Cox *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome. Res.* **10**, 1794 (2011).
34. C. Uttamapinant *et al.*, Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew. Chem. Int. Ed. Engl.* **51**, 5852 (2012).
35. W. Wen, J. L. Meinkoth, R. Y. Tsien, S. S. Taylor, Identification of a Signal for Rapid Export of Proteins from the Nucleus. *Cell* **82**, 463 (1995).
36. T. A. Brown *et al.*, Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction. *Mol. Cell. Biol.* **31**, 4994 (2011).
37. D. Kalderon, B. L. Roberts, W. D. Richardson, A. E. Smith, A Short Amino-Acid Sequence Able to Specify Nuclear Location. *Cell* **39**, 499 (1984).
38. E. L. Snapp *et al.*, Formation of stacked ER cisternae by low affinity protein interactions. *J. Cell. Biol.* **163**, 257 (2003).
39. A. Keppler, H. Pick, C. Arrivoli, H. Vogel, K. Johnsson, Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc.Natl. Acad. Sci. U.S.A.* **101**, 9955 (2004).
40. A. Chapman-Smith, J. E. Cronan, Jr., In vivo enzymatic protein biotinylation. *Biomol. Eng.* **16**, 119 (1999).
41. G. Kroemer, L. Galluzzi, C. Brenner, Mitochondrial membrane permeabilization in cell death. *Physiol. Rev.* **87**, 99 (2007).
42. W. R. Patterson, T. L. Poulos, Crystal structure of recombinant pea cytosolic ascorbate peroxidase. *Biochemistry* **34**, 4331 (1995).
43. K. H. Sharp, M. Mewies, P. C. Moody, E. L. Raven, Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nat. Struct. Biol.* **10**, 303 (2003).
44. M. Beck *et al.*, The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, (2011).

45. D. Bordo, P. Argos, Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J. Mol. Biol.* **217**, 721 (1991).
46. H. F. Becker, Y. Motorin, R. J. Planta, H. Grosjean, The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic. Acids. Res.* **25**, 4493 (1997).
47. K. Dresdner, L. A. Barker, M. Orlowski, S. Wilk, Subcellular distribution of prolyl endopeptidase and cation-sensitive neutral endopeptidase in rabbit brain. *J. Neurochem.* **38**, 1151 (1982).
48. F. N. Vogtle *et al.*, Intermembrane space proteome of yeast mitochondria. *Mol. Cell. Proteomics*. **11**, 1840 (2012).
49. M. Darshi *et al.*, ChChd3, an Inner Mitochondrial Membrane Protein, Is Essential for Maintaining Crista Integrity and Mitochondrial Function. *J. Biol. Chem.* **286**, 2918 (2011).
50. H. W. Chen *et al.*, Mammalian polynucleotide phosphorylase is an intermembrane space RNase that maintains mitochondrial homeostasis. *Mol. Cell. Biol.* **26**, 8475 (2006).
51. A. Zhyvoloup *et al.*, Subcellular localization and regulation of coenzyme a synthase. *J.Biol. Chem.* **278**, 50316 (2003).
52. M. Darshi, K. N. Trinh, A. N. Murphy, S. S. Taylor, Targeting and Import Mechanism of Coiled-coil Helix Coiled-coil Helix Domain-containing Protein 3 (ChChd3) into the Mitochondrial Intermembrane Space. *J. Biol. Chem.* **287**, 39480 (2012).
53. M. Sultan *et al.*, A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956 (2008).